



## Predicting The Target Genes For A Distal Regulatory Regions:

(Updated 24 October 2013, Mike Pazin)

The [Regulatory Elements Database](#) is an ENCODE-funded tool described in a recent [publication](#) that can be used to make predictions about the linkage between regulatory regions and genes, based on the statistical association of DNase I Hypersensitive Sites (DHS) and gene expression across more than 100 samples consisting of over 70 diverse cell types.

From the [Regulatory Elements Database](#) site, click on the “By COORDINATE” link ([Arrow 1](#)).

Regulatory Elements Database

This database provides a user interface to the results of the analysis presented in Sheffield et al. (2013), signal in 112 human samples. Questions? contact Nathan Sheffield. See also: [Supplemental Files](#)

There are 5 ways to explore the data:

- By CELLTYPE - Select cell-types to include or exclude.
- By CLUSTER - View promoter, CpG-islands, and conserved element overlap for all clusters, and select:
  - Muscle-specific cluster: cluster 1520
  - Prostate and hepatocyte cluster: cluster 910
  - Prostate-only cluster: cluster 2483
  - Hematopoietic cluster: cluster 25
  - Pluripotency cluster: cluster 104
- By GENE - Search by gene of interest
  - MyoG
  - RBFOX1
  - LIN28A
  - HBG1
- By COORDINATE - Give chr, start, stop to find all regulatory elements in a region:
  - IRF2 regulator: chr4: 185240845-185249995
  - MyoD1 regulator: chr11: 17828545-17828695
  - Blood regulator: chr3: 128166420-128166570
- By FACTOR - find a specific TF of interest
  - CTCF
  - Myf
  - AP1
  - Pou5f1 (Oct-4)

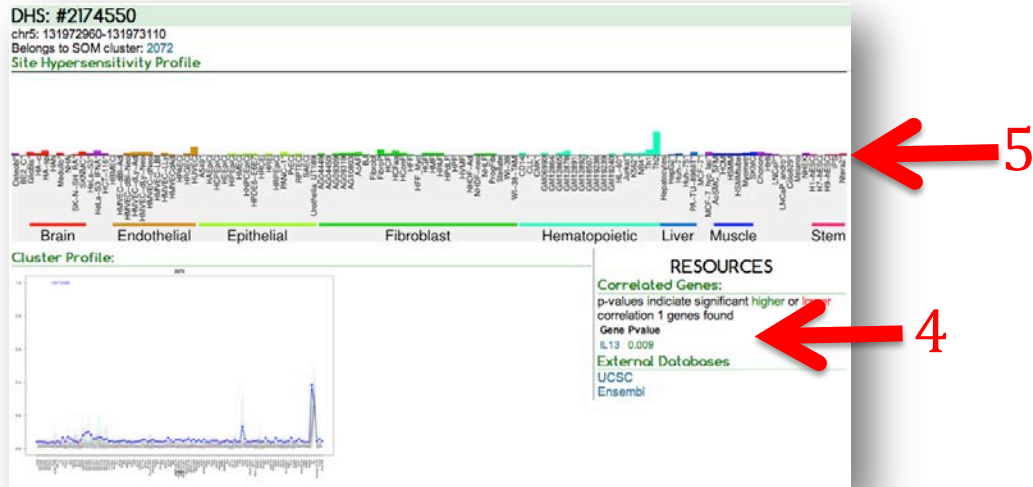
In the new screen, enter a genomic coordinate (such chr5:131960000-132000000) in the box, and click the “Submit” button. (For example, one could search  $\pm$  50 kb from a SNP of interest.)

Chromosome, start, stop:  Submit

5: 131960000-132000000  
23 DHS sites found

Site ID	Location	Cluster	Connections
2174568	chr5: 131988965-131989115	2001	(0) -
2174546	chr5: 131967940-131968090	2110	(1) - ENSG00000131437
2174547	chr5: 131968265-131968415	1605	(1) - ENSG00000169194
2174548	chr5: 131970980-131971130	2072	(1) - ENSG00000169194
2174550	chr5: 131972960-131973110	2072	(1) - ENSG00000169194
2174556	chr5: 131977240-131977390	2168	(1) - ENSG00000113520
2174557	chr5: 131977640-131977790	2072	(2) - ENSG00000169194, ENSG00000223442
2174560	chr5: 131982700-131982850	860	(1) - ENSG00000223442
2174564	chr5: 131985020-131985170	1605	(1) - ENSG00000169194
2174577	chr5: 131992180-131992330	1259	(1) - ENSG00000113520
2174578	chr5: 131992545-131992695	1078	(2) - ENSG00000223442, ENSG00000113520
2174582	chr5: 131993765-131993915	1970	(1) - ENSG00000113520
2174584	chr5: 131994205-131994355	1970	(1) - ENSG00000223442
2174587	chr5: 131995065-131995215	1704	(1) - ENSG00000113520
2174590	chr5: 131995740-131995890	369	(1) - ENSG00000113520
2174595	chr5: 131997040-131997190	715	(1) - ENSG00000113522
2174596	chr5: 131997225-131997375	514	(1) - ENSG00000113522
2174597	chr5: 131997380-131997530	514	(2) - ENSG00000113522, ENSG00000223442
2174600	chr5: 131998145-131998295	2072	(1) - ENSG00000113520
2174602	chr5: 131998645-131998795	218	(1) - ENSG00000113522
2174603	chr5: 131998800-131998950	72	(2) - ENSG00000113522, ENSG00000205089
2174605	chr5: 131999145-131999295	72	(2) - ENSG00000113522, ENSG00000205089
2174606	chr5: 131999680-131999830	715	(1) - ENSG00000113522

A list of the DHS found in the region is returned in a new screen. For each DHS, the location (genomic coordinate) is a hyperlink that can be used to view additional information (eg, **Arrow 2**). The heading “Connections” displays the Ensembl gene name(s) for the predicted target of each DHS (**Arrow 3**). The “Cluster” number groups DHS with a similar signal profile across cell types. Clicking on a DHS location (eg., **Arrow 2**) opens a new screen with additional details:



For the selected DHS, the display indicates predicted target gene(s) and p value(s) (**Arrow 4**), green for positive correlation with expression, red for negative correlation with expression. The DHS signal profile across individual cell types is also displayed (**arrow 5**).

In this example, this DHS found within the RAD50 gene is predicted to regulate the neighboring IL13 gene. [Publications indicate](#) the orthologous mouse DHS have been found to regulate the neighboring IL4, IL13, and IL5 genes, as predicted for the human DHS in this region.